SRI International

SCENE MODELING: A STRUCTURAL BASIS FOR IMAGE DESCRIPTION

Technical Note 221

July 1980

By: Jay M. Tenenbaum, Program Manager
Martin A. Fischler, Senior Computer Scientist
Harry G. Barrow, Senior Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

SRI Project 1019

The work reported herein was supported by the U. S. Army Research Office under Contract No. DAAG29-79-C-0216.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 1980		**00-07-1980 to 00-07-1980**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Scene Modeling: A Structural Basis for Image Description	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**30**	

SCENE MODELING: A STRUCTURAL BASIS FOR IMAGE DESCRIPTION

Jay M. Tenenbaum, Martin A. Fischler, and Harry G. Barrow

Artificial Intelligence Center, SRI International,
Menlo Park, California 94025

Conventional statistical approaches to image modeling are fundamentally limited because they take no account of the underlying physical structure of the scene nor of the image formation process. The image features being modeled are frequently artifacts of viewpoint and illumination that have no intrinsic significance for higher-level interpretation. In this paper a structural approach to modeling is argued for that explicitly relates image appearance to the scene characteristics from which it arose. After establishing the necessity for structural modeling in image analysis, a specific representation for scene structure is proposed and then a possible computational paradigm for recovering this description from an image is described.

I INTRODUCTION

Current research on image modeling appears dominated by attempts to characterize, mathematically and statistically, spatial variations of brightness in gray-level imagery. The basic premise underlying much of this work is that one can extract invariant pictorial features (such as regions of homogeneous brightness) that correspond to semantically meaningful entities (such as surfaces of objects).

It is our position that this abstract mathematical approach to image modeling has fundamental limitations because it takes no account of the underlying physical structure of the scene nor of the image formation process. The image features being modeled are thus frequently artifacts of viewpoint and illumination that have no intrinsic

significance for higher-level interpretation. To avoid artifacts, image appearance must be explicitly related to the scene structure from which it arose, primarily physical surface characteristics such as orientation, reflectance, color, and distance.

Models that represent image appearance in terms of physical scene structure will be called "structural models" to distinguish them from "statistical models" that restrict themselves to describing immediate image appearance.

The historical emphasis on statistical modeling, reflected in these proceedings, arose from two sources: the hope of finding simple solutions to image analysis problems which avoid the need to get involved with scene structure; and the fear that image formation is too complex a process to model deterministically. In this paper we will demonstrate that both the hope and fear are unfounded. We will first examine in detail the limitations of statistical models and build a strong case for structural models. We will then propose a specific representation for scene structure, called "intrinsic imagery," and describe a computational paradigm for recovering this description from an input image.

II STATISTICAL VERSUS STRUCTURAL COMPLEXITY

As a point of departure, it is important to reiterate that there are two distinct sources of image complexity: statistical and structural. Statistical complexity involves brightness variations that arise from truly random phenomena, such as noise or random dot textures [1]. Also included in this category are variations due to physical phenomena, such as atmospheric scatter, that are too complex to model in detail. Structural complexity, by contrast, refers to brightness variations that arise in a deterministic way from physical scene structure, such as surface albedo and orientation. While all real images contain some degree of statistical complexity, structural

complexity is often dominant, particularly for images of three-dimensional scenes.

Statistical image models are clearly appropriate for describing statistical complexity, but structural complexity requires structural models.

III NECESSITY FOR STRUCTURAL MODELS

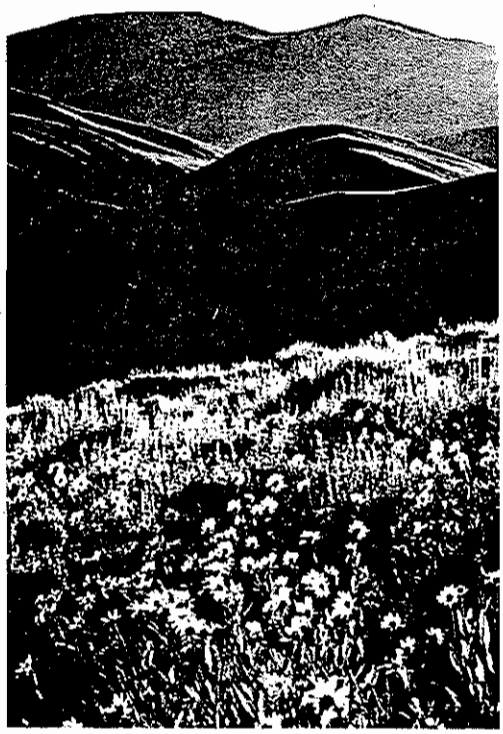
The need for structural models derives from three fundamental flaws in current statistical approaches: they make invalid statistical assumptions, they use ad hoc tests of statistical significance, and they produce impoverished descriptions having limited utility.

First, statistical models generally assume that an image is composed of regions of homogeneous brightness (or some statistic thereof), superimposed with random noise. Implicit is the assumption that such regions correspond to homogeneous surfaces in a scene. Real surfaces do, in fact, generally have uniform reflectance. Image brightness, however, depends on many factors besides surface reflectance, notably incident illumination and surface orientation. Consequently, homogeneous surfaces frequently appear in images as regions with high brightness gradient and even discontinuities, due to cast shadows. The geometry of imaging imposes similar artifacts on spatial properties of surfaces, such as the shape, size, and density of texture elements. Figure 1 contains striking examples of both phenomena.

The second limitation is perhaps more fundamental. Even if statistical homogeneity were a valid model (as is approximately the case for large areas of sky in Figure 1a), the problem of deciding what constitutes a statistically significant variation remains.

The traditional solution has been to attempt to remove known artifacts through a normalization process and then detect discontinuities using a threshold. While this philosophy is basically

(a) Varying incident illumination on different faces of the mountain transforms uniformly white snow into image regions with significantly varying hue and reflectance.



(b) The significant shading gradients on the background hills are again due to variations in angle of incident illumination rather than any intrinsic change in surface color. The sharp two-dimensional texture gradient (foreground) is an artifact of perspective, not a characteristic of flowers.

Despite the variations in image appearance, viewers correctly perceive the uniform reflectivity of snow and hills and the regular size and density of flowers.

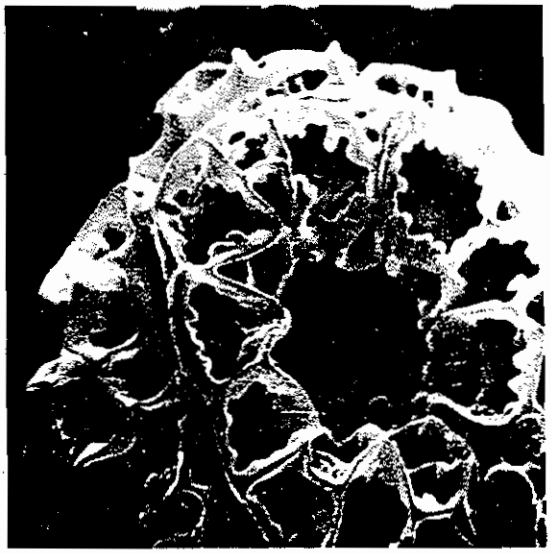
FIGURE 1 EXAMPLES REFUTING THE ASSUMPTION THAT HOMOGENEOUS SURFACE CHARACTERISTICS APPEAR AS HOMOGENEOUS IMAGE FEATURES

sound, implementations have failed in practice because they rely on ad hoc normalization schemes (such as histogram equalization). Normalization cannot be meaningfully accomplished without taking account of the causes of variation, which in turn depend on how the image is formed. Intensity variations and discontinuities resulting from illumination gradients, shadows, or a surface turning smoothly away from the illumination source are artifacts, no matter how large. Conversely, variations resulting from small step changes in surface reflectance or orientation can be very significant. Clearly, no purely statistical process can distinguish which image features correspond to significant scene events (i.e., surfaces and surface boundaries) and which do not (i.e., shadows and highlights).

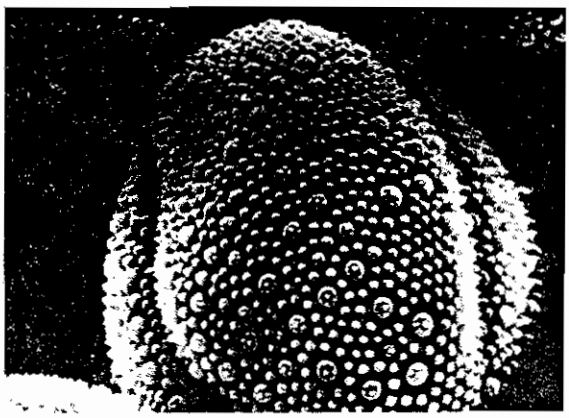
The third limitation of statistical modeling is the inadequacy of the resulting description for subsequent levels of interpretation. Even if an image could be reliably partitioned into homogeneous regions corresponding to surfaces of objects, the two-dimensional shape features used for region descriptions would still limit recognition to known objects observed from standard viewpoints; a three-dimensional characterization of surface shape is needed to recognize unfamiliar views of known objects and to assimilate multiple views of previously unseen objects.

We find it significant that when a human looks at images of scenes like those in Figure 1, what he sees, primarily, are the actual physical characteristics of three-dimensional surfaces, independent of artifacts of illumination and viewpoint. He tends to perceive color independent of illuminant, size independent of distance, and shape independent of orientation. Such constancies have been extensively validated in the psychological literature, and it is known that they do not depend on familiarity with scene content (see Figure 2). There is considerable evidence, in fact, that these physical normalizations are performed by "hardwired" neural circuitry, attesting to their fundamental importance; as Figure 3 illustrates, it is only with great effort that they can be overridden to expose the raw image content.

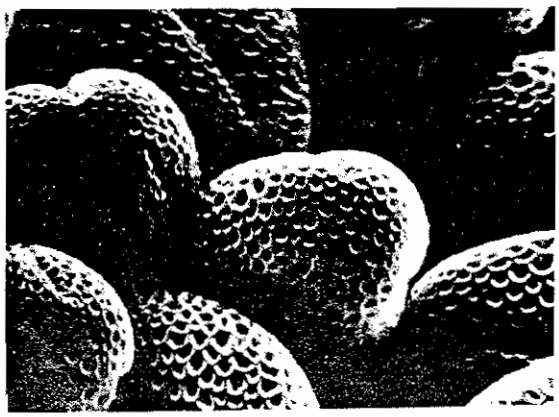
(a) Castanopsis (x 3500)



(b) Drimys (x 3200)

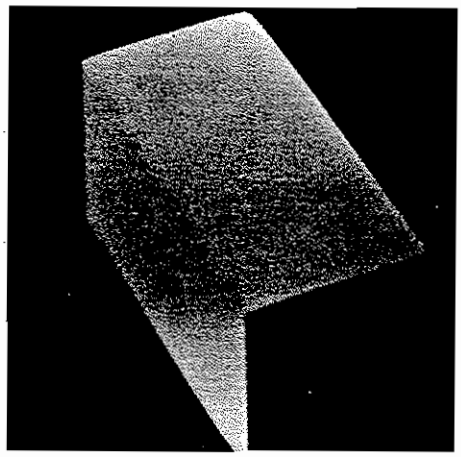


(c) Flax (x 1000)

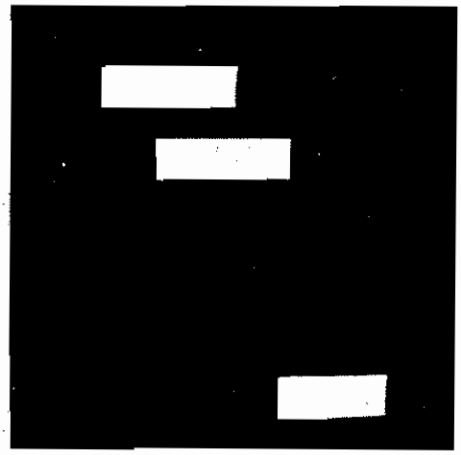


(d) Wallflower (x 1800)

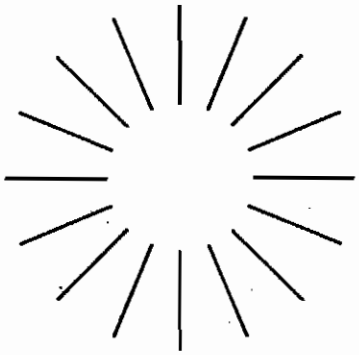
Figure 2 Photomicrographs of Pollen Grains (Macleod [2])



3a



3b



3c

3a - Low-contrast interior boundary 3b - Masked cross sections of the interior boundary in Figure 3a 3c - A subjective contour

In Figure 3a, an intersection boundary is clearly perceived. Yet when that area of the image is viewed through a mask that exposes only local cross sections, no local contrast is visible (Figure 3b). The boundary perceived in Figure 3a is demanded by the integrity of the surfaces it joins. Subjective contour illusions, like the so-called sun illusion (Figure 3c), appear to be an extreme example of this same phenomenon where an edge is clearly perceived despite the complete absence of local evidence. A plausible explanation is that the edge corresponds to the boundary of an occluding disk-shaped surface, whose presence is implied by the abrupt line endings [3].

Figure 3. Demonstration that Human Perception of Surface Boundaries is Not Critically Dependent on Image Contrast

In summary, we believe that the obfuscation of 3-D scene structure by 2-D image features is a fundamental limitation of current image analysis systems and an important reason why their performance is so inferior to that of the human visual system.

IV A PARADIGM FOR STRUCTURAL MODELING

Having argued for the necessity for structural modeling, we must now establish what scene characteristics to model, how to represent them in a computer, and how to recover a structural description from an image.

A. Intrinsic Characteristics and their Representation

When an image is formed, whether by eye or camera, the light intensity at each point in the image is determined by three main factors at the corresponding point in the scene: the incident illumination, the surface reflectance, and the surface orientation. In the simple case of an ideally diffusing surface, for example, the image light intensity L is given by Lambert's law:

$$L=I*R*COS\ i \qquad [1]$$

where I is incident illumination, R is surface reflectance (albedo), and i is angle of incidence with respect to the local surface normal.

In more complicated scenes, additional factors such as specularity, transparency, luminosity, distance, and so forth must be considered. We call these properties intrinsic characteristics to distinguish them from image features which have no physical significance.

A suitable representation for intrinsic characteristics, consistent with their acquisition and subsequent use, is a set of iconic arrays in registration with the original image array. Each array contains values for a particular characteristic of the surface element visible at the corresponding point in the sensed image. Each array also contains

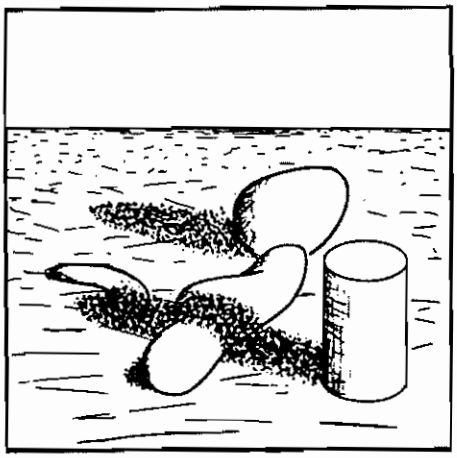
explicit indications of boundaries due to discontinuities in value or gradient. We call such arrays "intrinsic images." Figure 4 gives an example of one possible set of intrinsic images corresponding to a monochrome image of a simple scene.

A concrete example of intrinsic images and their usefulness in computer vision can be seen in Figure 5, which summarizes experiments by Nitzan, Brain, and Duda with a scanning laser range finder [4]. This instrument directly measures the intrinsic properties of distance and apparent reflectance, based on the phase and amplitude of the signal received as a modulated laser beam is scanned over the scene.

The left side of Figure 5b is a range image of the scene in Figure 5a, with brightness inversely related to range. Note the absence of surface markings on the top of the cart and on the gray-level test chart. The right side of Figure 5b is a reflectance image, obtained by measuring returned amplitude and using the range to compensate for varying angles of incidence. Note the absence of shadows and shading gradients present in the original intensity image (Figure 5a).

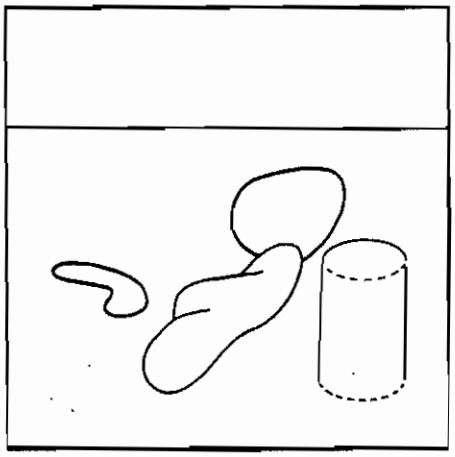
Because the range data are uncorrupted by reflectance variations and the amplitude data are unaffected by ambient lighting and shadows, it is easy to extract surfaces of uniform height (Figure 5c) or reflectivity (Figure 5e) and surface boundaries where range is discontinuous (Figure 5d). Such tasks are difficult to perform reliably in gray-level imagery; but with pure range and amplitude data, even simple-minded techniques such as thresholding and region-growing work well. In intrinsic images, the assumptions underlying statistical image modeling are valid!

Instrumentation such as a laser range finder trivializes the problem of extracting intrinsic surface characteristics from sensory data. A key question is whether such information can be recovered from a single gray-level image, which is all that is available in many image analysis applications. While no definitive answer can yet be given, human performance and recent computer vision research both give cause for optimism. In following sections we discuss the computational

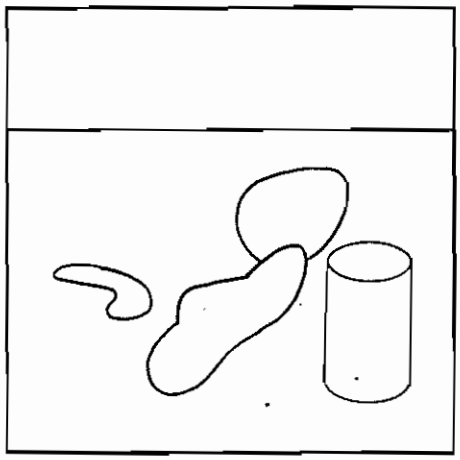


(a) ORIGINAL SCENE

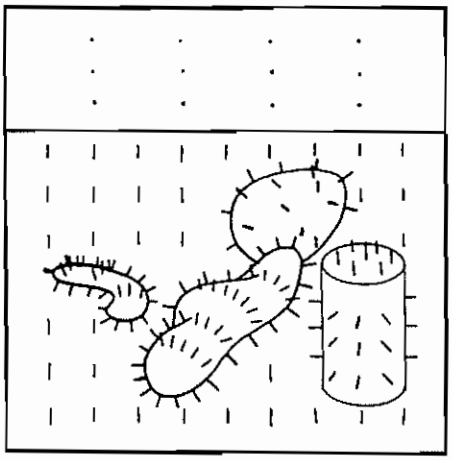
The images are depicted as line drawings, but, in fact, would contain values at every point. The solid lines in the intrinsic images represent discontinuities in the scene characteric; the dashed lines represent discontinuities in its derivative. In the input image, intensities correspond to the reflected light flux received from the visible points in the scene. The distance image gives the range along the line of sight from the center of projection to each visible point in the scene. The orientation image gives a vector representing the direction of the surface normal at each point. The reflectance image gives the albedo (the ratio of total reflected to total incident illumination) at each point.



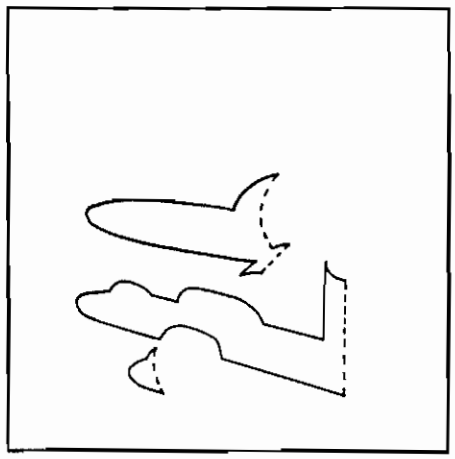
(b) DISTANCE



(c) REFLECTANCE

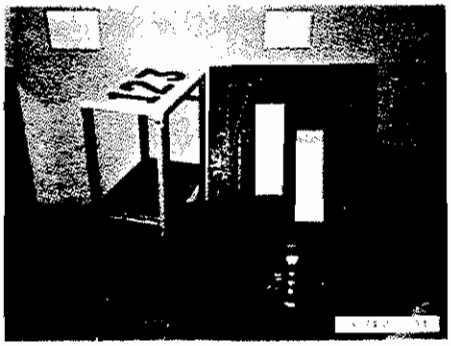


(d) ORIENTATION (VECTOR)

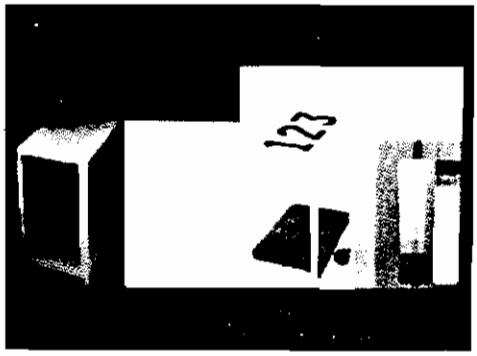


(e) ILLUMINATION

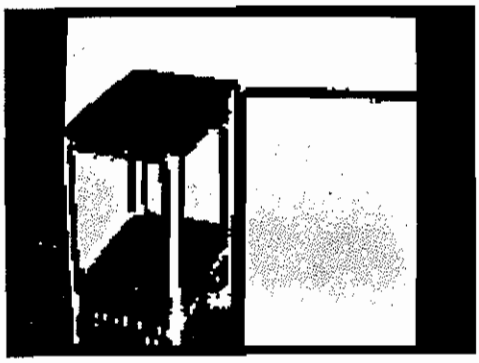
Figure 4 A Set of Intrinsic Images Derived from a Single Monochrome Intensity Image



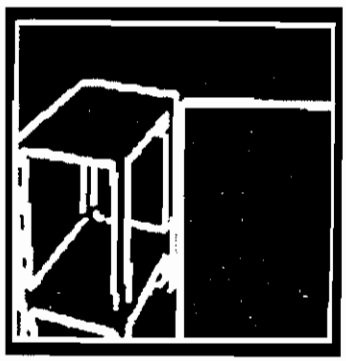
(a) A CONVENTIONAL PHOTO OF A SCENE



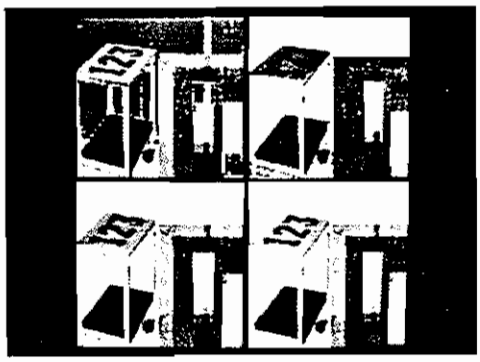
(b) DISTANCE AND REFLECTANCE IMAGES



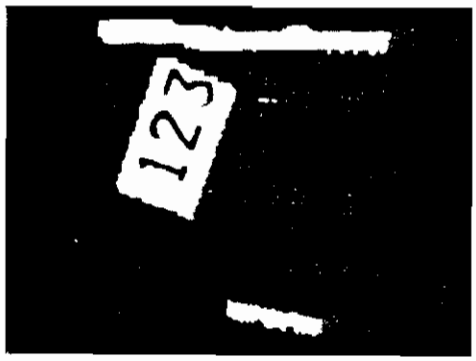
(c) EXTRACTED PLANAR SURFACES



(d) DISCONTINUITIES IN RANGE



(e) THRESHOLDING REFLECTANCE



(f) CORRECTED VIEW OF CART TOP

Figure 5 Experiments with a Laser Range Finder

problems involved in recovering intrinsic characteristics and briefly review promising solutions that have been proposed and demonstrated for the case of a single gray-level image. Research is also proceeding at SRI International and elsewhere for less deprived cases, such as where stereopsis [5], motion parallax [6], or a priori object models are available [7].

B. Recovery of Intrinsic Characteristics

The central problem in recovering intrinsic characteristics is that the desired information is confounded in the sensory data. Photometrically, the light intensity observed at each point in an image can result from any of an infinitude of illumination, reflectance, and orientation combinations at the corresponding scene point (see, for example, Equation 1). Geometrically, each point in the image can correspond to any point along a ray in space (see Figure 6). Recovery is thus an underconstrained problem that requires additional constraints for solution.

The necessary constraints follow from assumptions about the nature of the scene being viewed and the physics of the imaging process. In images of three-dimensional scenes, the brightness values are not independent but are constrained by various physical phenomena. Since surfaces are continuous in space, their characteristics (reflectance, orientation, range) are generally continuous across an image, except at surface boundaries. Incident illumination also usually varies smoothly over a scene, except at shadow boundaries. Elementary photometry tells us that where all intrinsic characteristics are continuous, image brightness is continuous; conversely, where one or more intrinsic characteristics are discontinuous, a brightness edge will usually result. The pattern of brightness variation in an image can provide important clues as to the local behavior of the intrinsic characteristics.

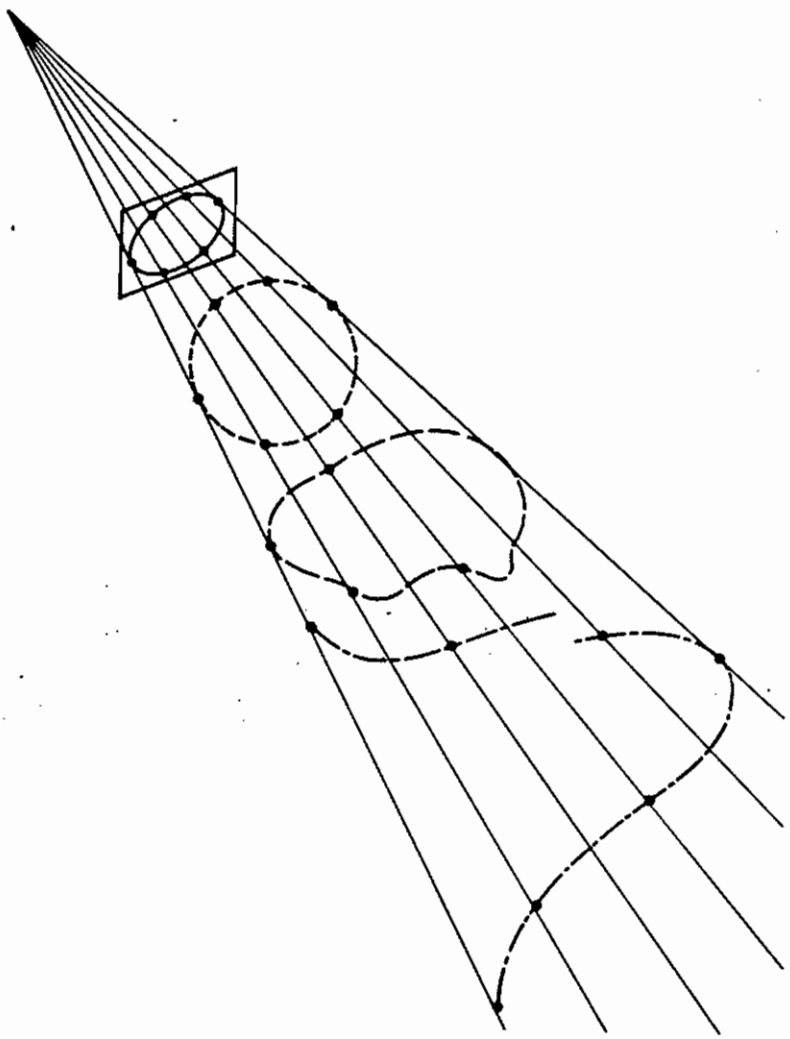


Figure 6 Three-Dimensional Conformation of Lines Depicted in a Line Drawing is Inherently Ambiguous

1. Shape from Shading

Horn showed that shading variations could be used to determine the three-dimensional shape of a Lambertian surface with uniform reflectance, viewed under distant point-source illumination [8 and 9]. The basic approach can be grasped from Equation 2, which was obtained from Equation 1 by taking logs and differentiating.

$$dL/L = dI/I + dR/R + d(\text{Cos } i)/\text{Cos } i \qquad [2]$$

Under Horn's assumptions, illumination and reflectance are both uniform, so that percentage changes in image brightness are directly proportional to percentage changes in (the cosine of) orientation, as shown in Equation 3.

$$dL/L = d(\text{Cos } i)/\text{Cos } i \qquad [3]$$

Horn showed that given suitable boundary conditions, Equation 3 can be integrated to recover shape.

2. Lightness from Shading

A second example of exploiting photometry, also due to Horn [10], involves the recovery of surface lightness (a psychological term representing relative reflectance). This study assumed a planar surface, viewed under smoothly varying illumination. The surface was painted with a patchwork of contrasting regions, each of uniform reflectance. Within regions, since orientation and reflectance are both constant, variations in image brightness are directly proportional to variations in illumination, as given by Equation 4a. Moreover, since illumination was assumed to vary smoothly, the brightness gradient within regions is small. At region boundaries, however, reflectance jumps discontinuously and dominates the small illumination gradient, as expressed in Equation 4b. To recover surface lightness, Horn first spatially differentiated the input (log brightness) image to obtain a gradient image and then thresholded it to eliminate slow illumination

variations. The remaining discontinuities, assumed to be reflectance jumps at the edges of regions (as in Equation 4b) were then reintegrated to recover the relative lightnesses of the various regions.

$$dL/L = dI/I \quad \text{(within regions)} \qquad [4a]$$

$$dL/L = dR/R \quad \text{(at boundaries)} \qquad [4b]$$

Surface color independent of illuminant can be estimated from lightness values recovered independently in three spectral bands, analogous to Land's retinex theory [11].

3. Shape From Contour

Horn's work emphasized photometric cues, but geometric cues to 3-D surface structure are at least as valuable (see Figure 7).

The ability to perceive surface structure from line drawings is truly remarkable since, as Figure 6 showed, each two-dimensional image curve can, in principle, correspond to an infinitude of possible three-dimensional space curves. However, people are not aware of this massive ambiguity. For example, when asked to provide a three-dimensional interpretation of an ellipse, the overwhelming response is a tilted circle, not some bizarrely twisting (or even discontinuous) curve that has the same image. As in Horn's work, some a priori assumptions about the scene must again be invoked. Recent research at SRI [12] and MIT [13 and 14] suggests that humans resolve the projective ambiguity by perceiving the smoothest possible space curve corresponding to a given image curve. Mathematically, they seek the space curve having the most uniform curvature and the least torsion, as expressed by minimizing the terms in Equation 5.

$$\int \left(\frac{dkb}{ds}\right)^2 ds = \int (\dot{k}^2 + k^2 t^2) ds \qquad [5]$$

(Here k is the local differential curvature, k' is its spatial derivative along the curve, B is the binormal, and t is the torsion.)

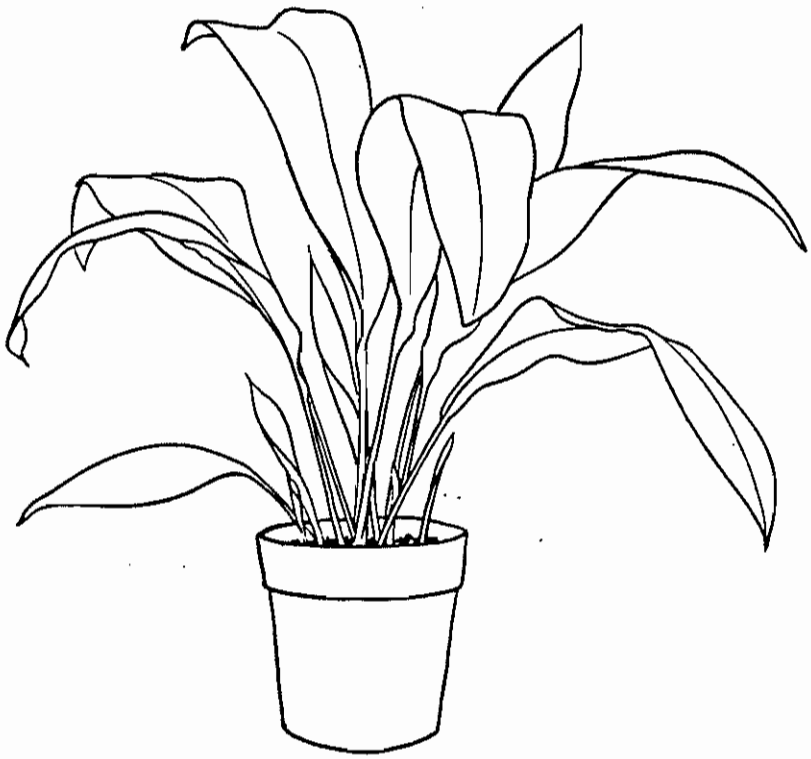


Figure 7 Line Drawing of a Three-Dimensional Scene

The smoothness assumption expressed by Equation 5 has both an ecological and a statistical rationale. Ecologically, assumptions on smoothness of curves follow from the earlier assumptions on smoothness of surfaces, which, in turn, are rooted in assumptions that physical surfaces assume minimum energy configurations. Statistically, it is reasonable to assume that a scene is being viewed from a general position so that perceived smoothness is not an accident of viewpoint. In Figure 6, for example, the discontinuous curve projects into an ellipse from only one viewpoint. Thus such a curve would be a highly improbable three-dimensional interpretation of an ellipse.

A computer program has been written, based on Equation 5, that can successfully determine three-dimensional space curves corresponding to simple image curves. Referring to Figure 8, points along an image curve define rays in space along which the corresponding space curve points are constrained to lie. The program can adjust the distance associated with each space curve point by sliding it along the ray like a bead on a wire. An iterative optimization procedure determines the configuration of points that minimize the integral in Equation 5. Optimization proceeds by independently adjusting each space curve point and observing the incremental change in local curvature and torsion. (Note that local perturbations have only local effects.) Witkin [13] used a similar approach to model the perception of planar orientation associated with simple closed curves.

The program produces correct 3-D interpretations for simple open and closed curves, such as interpreting an ellipse as a tilted circle and a trapezoid as a tilted rectangle. However, convergence is slow and somewhat dependent on the initial choice of z-values.

4. Shape from Texture

Texture gradient is a well-known geometric clue for inferring three-dimensional surface structure. The variations of size, density, eccentricity, and orientation of the texture elements in Figure 9 are hardly random; they are predicable consequences of the foreshortening

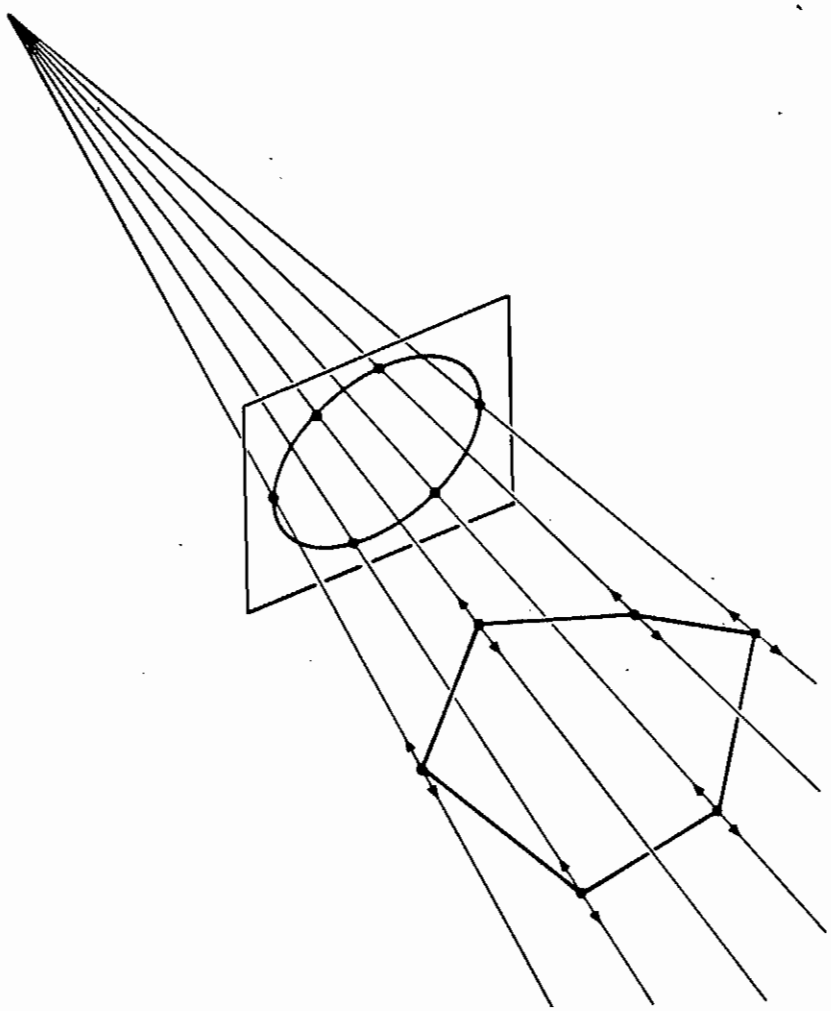


Figure 8 An Iterative Procedure for Determining the Optimal Space Curve Corresponding to a Given Line Drawing

that occurs when a tilted surface is imaged under perspective projection. A recent thesis by Stevens [14] provides mathematical formulations for a number of texture depth cues, which previously have been described only qualitatively by psychologists. Another recent thesis by Kender establishes the principle that textured surfaces are perceived at orientations that maximize the regularity, homogeneity, and symmetry of the texture [15].

Underlying Kender's work on texture and our own work on line drawings is a fundamental assumption that the world is generally isotropic. This assumption, which we call "generalized isotropy," is reminiscent of the Gestalt notions of Praegnanz, but mathematically more precise.

5. The Role of Edges

The previous sections described various ways in which surface characteristics could be inferred from image features, such as shading, contour, and texture. However, in each case the physical nature of the image feature being interpreted was known. For example, in determining shape from shading, Horn assumed that only surface orientation was changing; in determining lightness, he assumed that only reflectance was discontinuous; in determining shape from contour, we implicitly assumed that image curves corresponded to surface boundaries and were not shadows or lines painted on a flat surface. When a scene contains many occluding objects that may be varicolored or cast shadows, such simple assumptions are invalid. One is then faced with the problem of deciding what physical characteristic (or characteristics) is, in fact, responsible for an observed intensity variation and which characteristics are discontinuous across intensity edges.

The pattern of brightness variation on either side of an intensity edge can sometimes provide strong clues as to the type of scene event responsible (shadow or surface boundary), and thus to which intrinsic characteristics are actually discontinuous at that point. A simple example is the continuity of texture and high contrast at shadow

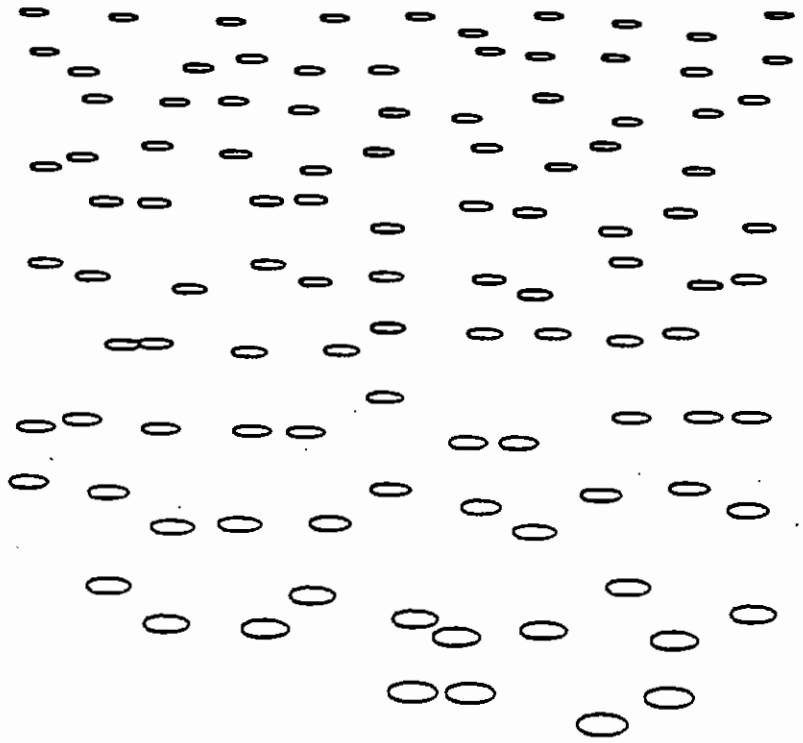


Figure 9 Texture Gradients on a Tilted Planar Surface (Stevens [14])

edges, indicating a discontinuity in illumination. The interpretation of brightness edges as scene events is also important because knowing the type of scene event sometimes allows explicit values to be determined for some of the intrinsic characteristics. For example, at an extremal occluding boundary, where an object curves smoothly away from the viewer, the surface orientation can be inferred exactly at every point along the boundary. (The local surface normal is constrained to be normal to the edge element in the image and normal to the line of sight.) A test for extremal boundaries can be made by determining whether the observed brightness variation along an edge is consistent with assumed extremal orientations [3].

Having identified a number of edge and surface constraints, it was necessary to establish whether they were sufficient to permit the simultaneous recovery of surface reflectance and orientation from a single image. We defined the simplest domain in which confounding problems arose in a general way and proceeded to exhaustively catalog the physical interpretations corresponding to all possible image intensity patterns [3]. The domain consisted of smooth (no creases, folds), uniformly reflecting Lambertian surfaces, illuminated by a distant point source, and uniform, diffuse background light (approximating sun and sky). The resulting catalog is reproduced in Table 1. With but two exceptions, brightness discontinuities could be unambiguously interpreted as physical events. Furthermore, in most cases, values for one or more intrinsic characteristics were either determined or strongly constrained.

We concluded that the recovery problem was mathematically well posed, at least for the simple domain. The clues and constraints obtained by interpreting image discontinuities according to the catalog define a system of equations and inequalities, relating the intensity values to the intrinsic characteristics and boundary conditions at the discontinuities. In principle, this system can be solved, yielding feasible values for the intrinsic characteristics. In practice, however, the equations are highly nonlinear and boundary conditions may

Table 1

The Nature of Edges

LA and LB refer to variations of intensity along sides A and B of an edge. Intensities are either constant, varying, or varying in accordance with the assumed orientations along an extremal boundary, the so-called tangency condition.

Region Intensities LA	LB	Edge Type	Region Types	Intrinsic Edges Intrinsic Values D	N	R	I
Constant	Constant	Occluding sense unknown	A B shadowed	EDGE	EDGE	EDGE RA RB	IA IB
Constant	Varying	1 Shadow	A shadowed B illuminated		NB.S	RA RB	EDGE IA IB
		2 A occludes B	A shadowed B illuminated	EDGE DA<DB	EDGE NA	EDGE RA	EDGE IA
Varying	Varying	Inconsistent with domain					
Constant	Tangency	B occludes A	A shadowed B illuminated	EDGE DA>DB	EDGE NB	EDGE RA RB	EDGE IA IB
Varying	Tangency	B occludes A	A B illuminated	EDGE DA>DB	EDGE NB	EDGE RB	EDGE IB IA
Tangency	Tangency	Not seen from general position					

not always be determined reliably. This suggested an iterative numerical solution process, such as relaxation.

6. A Computational Model

A parallel computational model was proposed to illustrate how recovery might be performed. The basic model, reproduced in Figure 10, can be regarded as a generalization of Horn's lightness model [10] and Marr and Poggio's cooperative stereopsis model [5], that simultaneously recovers both geometric and photometric attributes. In essence, it consisted of a stack of registered arrays representing the original intensity image (top) and the primary intrinsic image arrays. Processing was initialized by detecting intensity edges in the original image, interpreting them according to the catalog of appearances, and then creating the appropriate edges in the intrinsic images (as implied by the descending arrows).

Parallel local operations (shown as circles) modified the values in each intrinsic image to make them consistent with intraimage continuity and limit constraints (for example, reflectance must be between 0 and 1). Simultaneously, a second set of processes (shown as vertical lines) operated to make the values at each point consistent with the corresponding intensity value, as required by the interimage photometric constraint. A third set of processes (shown as Xs) operated to insert and delete edge elements, which inhibit continuity constraints locally. The constraint and edge modification processes operated continuously and interacted to recover accurate intrinsic scene characteristics and to perfect the initial edge interpretation.

The action was envisaged to resemble an analog computer: as the value in one image increased, corresponding values in other images increased or decreased to maintain consistency with the observed intensity at that point. Within each image, values tended to propagate in from boundary conditions established along edges. This resembles relaxation processes used in physics for determining temperature or potential over a region from boundary conditions.

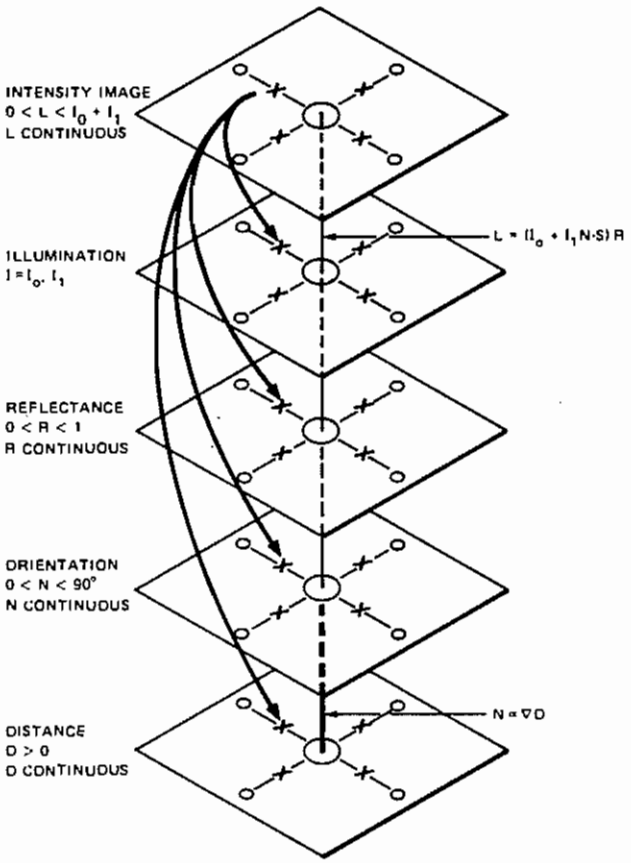


Figure 10 A Parallel Computational Model
for Recovering Intrinsic Images

How well does the proposed model work in the simplified domain? In theory exact recovery is generally possible for nonshadowed regions; where it is not possible, because of inadequate information in the original image, plausible estimates can usually be obtained. These results were significant, despite the simplicity of the domain, because they demonstrated for the first time the theoretical possibility of simultaneously recovering orientation, reflectance, and illumination from a single monochrome image, without recourse either to object prototypes or to primary depth cues, such as stereopsis, motion parallax, or texture gradient. (Such additional cues can, of course, be added to aid initialization in shadowed areas.)

We are currently implementing a version of the model and will soon test it on synthesized images of scenes from the simple domain. We are simultaneously studying the extensibility of the approach to more complex domains. We are convinced that the model can be extended to handle objects with creases, folds, painted-surface markings (texture), and other complicating features. Coping with more complex illuminations, however, appears difficult.

The theory of recovery outlined above depends heavily on analytic photometry and a precise lighting model, both for edge classification and for inferring shape from shading. Unfortunately, surfaces in real scenes have complex reflectance functions that are often strongly directional. Illumination patterns are equally complex, encompassing such phenomena as shading gradients from nearby or extended sources and secondary illumination by light reflected from nearby specular surfaces. The use of analytic photometry appears very questionable under these conditions. We are therefore working on an alternative theory of recovery that relies primarily on geometric cues, such as contour and texture gradient for gross shape determination, and uses photometry, only in a qualitative way, to indicate subtle refinements (e.g., bumps and dents).

V DISCUSSION

In this paper we have argued that the explicit modeling of scene structure is a critical first step in the interpretation of images of three-dimensional scenes and provided a demonstration that, theoretically, such information can be obtained even from a single gray-level image. More generally, in almost any image analysis task, an understanding of the relationship between scene content and image appearance is necessary for meaningful interpretation. Such relationships invariably have an underlying physical basis.

We are not attempting to denigrate the role played by statistical techniques in image analysis. Clearly some images are inherently statistical and contain little structure. More generally, because images are noisy and ambiguous, their interpretation requires fitting a priori models to the observed data; interpretation is thus a proper subset of statistical decision theory. The problem with conventional statistical approaches is that in ignoring the physical nature of the scene and the imaging process, they are forced to base interpretation on ad hoc and often invalid assumptions. In the paradigm we are suggesting, structural scene models provide a rational basis for decision-making.

In conclusion, while both statistical and structural models play important roles in image analysis, whenever image variation can be accounted for in structural terms, there are compelling reasons for doing so.

VI ACKNOWLEDGEMENTS

The work reported in this paper was performed under SRI's research program in computational vision, which is supported by ARPA, NSF, and NASA.

REFERENCES

1. B. Julesz, "Experiments in the Visual Perception of Texture," Scientific American, Vol. 232, pp. 34-43 (April 1975).

2. I.D.G. Macleod, "A Study in Automatic Photo Interpretation," Ph.D. Thesis, Dept. of Engineering Physics, Australian National University, Canberra, Australia (1970).

3. H. G. Barrow and J. M. Tenenbaum, "Recovering Intrinsic Scene Characteristics from Images," in Computer Vision Systems, A. Hanson and E. Riseman, eds., pp. 3-26 (Academic Press, New York, New York, 1978).

4. D. Nitzan, A. E. Brain, and R. O. Duda, "The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis," Proc. IEEE, Vol. 65, No. 2, pp. 206-220 (1977).

5. D. Marr and T. Poggio, "Cooperative Computation of Stereo Disparity," Science, Vol. 194, pp. 283-287 (1977).

6. S. Ullman, The Interpretation of Visual Motion (MIT Press, Cambridge, Massachusetts, 1979).

7. G. Falk, "Interpretation of Imperfect Line Data as a Three-Dimensional Scene," Artificial Intelligence, Vol. 4, No. 2, pp. 101-144 (1972).

8. B.K.P. Horn, "Obtaining Shape from Shading Information," in The Psychology of Computer Vision, P. H. Winston, ed. (McGraw-Hill, New York, New York, 1975).

9. B.K.P. Horn, "Understanding Image Intensities," Artificial Intelligence, Vol. 8, No. 2, pp. 201-231 (1977).

10. B.K.P. Horn, "Determining Lightness from an Image," Computer Graphics and Image Processing, Vol. 3, pp. 277-299 (1974).

11. E. H. Land, "The Retinex Theory of Color Vision," Scientific American, Vol. 237, No. 6, pp. 108-128 (December 1977).

12. H. G. Barrow and J. M. Tenenbaum, "Recovery of 3-D Shape Information from Image Boundaries" (in preparation).

13. A. Witkin, "The Minimum Curvature Assumption and Perceived Surface Orientation," presentation at Optical Society of America, November 1978.

14. K. Stevens, "Surface Perception from Local Analysis of Texture and Contour," Ph.D. Dissertation, Electrical Engineering and Computer

Science, Massachusetts Institute of Technology, Cambridge, Massachusetts (February 1979).

15. J. R. Kender, "Shape from Texture: A Computational Paradigm," Proc. ARPA IU Workshop, pp. 134-138, Palo Alto, California (April 1979).